Unidad I: Teoría de Aproximación.

José Luis Ramírez B.

November 11, 2024

Motivación.

- La gran mayoría de los modelos matemáticos que describen procesos físicos no pueden resolverse analíticamente.

Motivación.

- La gran mayoría de los modelos matemáticos que describen procesos físicos no pueden resolverse analíticamente.
- En una situación práctica, un problema matemático deriva de un fenómeno físico sobre el cual se han hecho algunas suposiciones para simplificarlo y poderlo representar matemáticamente.

Motivación.

- La gran mayoría de los modelos matemáticos que describen procesos físicos no pueden resolverse analíticamente.
- En una situación práctica, un problema matemático deriva de un fenómeno físico sobre el cual se han hecho algunas suposiciones para simplificarlo y poderlo representar matemáticamente.
- Una vez formulado el problema, deben diseñarse métodos numéricos para resolver el problema. La selección o construcción de los algoritmos apropiados cae propiamente dentro del terreno del Análisis Numérico.

Teoría de Errores y Aproximación.

El análisis numérico proporciona métodos computacionales para el estudio y solución de problemas matemáticos. Debido a que muchos cálculos son realizados en computadores digitales, es conveniente la discusión para la implementación de los métodos numéricos como programas de computador..

Teoría de Errores y Aproximación.

- La aparición de computadores ha hecho posible la solución de problemas, que por su tamaño antes eran excluidos.

Teoría de Errores y Aproximación.

- La aparición de computadores ha hecho posible la solución de problemas, que por su tamaño antes eran excluidos.
- Desafortunadamente los resultados son afectados por el uso de la Aritmética de Precisión Finita.

Teoría de Errores y Aproximación.

- La aparición de computadores ha hecho posible la solución de problemas, que por su tamaño antes eran excluidos.
- Desafortunadamente los resultados son afectados por el uso de la Aritmética de Precisión Finita.
- Esperamos tener siempre expresiones verdaderas como $2+2=4$, $3^2=9$, $(\sqrt{5})^2=5$, pero en la aritmética de precisión finita $\sqrt{5}$ no tiene un solo número fijo y finito, que lo representa.

Teoría de Errores y Aproximación.

- La aparición de computadores ha hecho posible la solución de problemas, que por su tamaño antes eran excluidos.
- Desafortunadamente los resultados son afectados por el uso de la Aritmética de Precisión Finita.
- Esperamos tener siempre expresiones verdaderas como $2+2=4$, $3^2=9$, $(\sqrt{5})^2=5$, pero en la aritmética de precisión finita $\sqrt{5}$ no tiene un solo número fijo y finito, que lo representa.
- En el computador se le da un valor aproximado cuyo cuadrado no es exactamente 5, aunque con toda probabilidad estará lo bastante cerca a él para que sea aceptable.

Teoría de Errores y Aproximación.

Un método numérico es un procedimiento mediante el cual se obtiene, de manera aproximada, la solución de ciertos problemas. Los resultados numéricos están influenciados por muchos tipos de errores, los cuales pueden ser catalogados a grandes rasgos en tres tipos básicos:

Teoría de Errores y Aproximación.

Un método numérico es un procedimiento mediante el cual se obtiene, de manera aproximada, la solución de ciertos problemas. Los resultados numéricos están influenciados por muchos tipos de errores, los cuales pueden ser catalogados a grandes rasgos en tres tipos básicos:

- **Errores inherentes** que existen en los valores de los datos de entrada, ya sea causados por incertidumbre o por la naturaleza necesariamente aproximada de la representación.

Teoría de Errores y Aproximación.

Un método numérico es un procedimiento mediante el cual se obtiene, de manera aproximada, la solución de ciertos problemas. Los resultados numéricos están influenciados por muchos tipos de errores, los cuales pueden ser catalogados a grandes rasgos en tres tipos básicos:

- **Errores inherentes** que existen en los valores de los datos de entrada, ya sea causados por incertidumbre o por la naturaleza necesariamente aproximada de la representación.
- **Errores de discretización** (llamados también de truncamiento) que surgen al reemplazar procesos límites por su resultado antes de alcanzar tal límite.

Teoría de Errores y Aproximación.

Un método numérico es un procedimiento mediante el cual se obtiene, de manera aproximada, la solución de ciertos problemas. Los resultados numéricos están influenciados por muchos tipos de errores, los cuales pueden ser catalogados a grandes rasgos en tres tipos básicos:

- **Errores inherentes** que existen en los valores de los datos de entrada, ya sea causados por incertidumbre o por la naturaleza necesariamente aproximada de la representación.
- **Errores de discretización** (llamados también de truncamiento) que surgen al reemplazar procesos límites por su resultado antes de alcanzar tal límite.
- **Errores de redondeo** que se originan al utilizar una aritmética que involucra números con un número finito de dígitos.

Errores Absolutos y Relativos.

Sea x el valor exacto de un número real y $\tilde{x}$ el valor aproximado. Contemplando todos los posibles errores, la relación entre el resultado exacto y el aproximado es:

$$x = \tilde{x} + E$$

Errores Absolutos y Relativos.

Sea x el valor exacto de un número real y $\tilde{x}$ el valor aproximado. Contemplando todos los posibles errores, la relación entre el resultado exacto y el aproximado es:

$$x = \tilde{x} + E$$

Se define el error absoluto y se denota E_a como la diferencia $x - \tilde{x}$, y se expresa siempre en valor absoluto.

$$E_a = |x - \tilde{x}|$$

Forward and Backward Error

- Sea x un número real y $f : \mathbb{R} \to \mathbb{R}$ una función. Si $\tilde{y}$ es un número real que es una aproximación a $y = f(x)$, entonces el error hacia adelante (Forward) en $\tilde{y}$ es la diferencia $\Delta y = \tilde{y} - y$.

Forward and Backward Error

- Sea x un número real y $f : \mathbb{R} \to \mathbb{R}$ una función. Si $\tilde{y}$ es un número real que es una aproximación a $y = f(x)$, entonces el error hacia adelante (Forward) en $\tilde{y}$ es la diferencia $\Delta y = \tilde{y} - y$.
- Sea $x \in \mathbb{R}$ y $f : \mathbb{R} \to \mathbb{R}$ una función. Supóngase que $\tilde{y}$ es una aproximación a $y = f(x)$ y $\tilde{y}$ está en el rango de f, es decir, $\tilde{y} = f(\tilde{x})$ para algún $\tilde{x}$, entonces la cantidad $\Delta x = \tilde{x} - x$ es el error hacia atrás (Backward) en $\tilde{y}$.

Forward and Backward Error

Ejemplo:

Supóngase que se desea calcular $y = \sqrt{2}$ y se obtiene $\tilde{y} = 1.4$, entonces:

Forward and Backward Error

Ejemplo:

Supóngase que se desea calcular $y = \sqrt{2}$ y se obtiene $\tilde{y} = 1.4$, entonces:

- Forward Error:
  $|\Delta y| = |\tilde{y} - y| = |1.4 - 1.4142\ldots| \approx 0.0142\ldots$

Forward and Backward Error

Ejemplo:

Supóngase que se desea calcular $y = \sqrt{2}$ y se obtiene $\tilde{y} = 1.4$, entonces:

- Forward Error:
  $|\Delta y| = |\tilde{y} - y| = |1.4 - 1.4142\ldots| \approx 0.0142\ldots$
- Backward Error: Nótese que $\sqrt{1.96} = 1.4$, entonces
  $|\Delta x| = |\tilde{x} - x| = |1.96 - 2| = 0.04$

Errores Absolutos y Relativos.

Una debilidad de esta definición es que la magnitud del error verdadero depende de la escala.

Errores Absolutos y Relativos.

Una debilidad de esta definición es que la magnitud del error verdadero depende de la escala.

- Por ejemplo podemos medir una barra en centímetros o en metros. Si la longitud exacta de la barra es $1m$ y por la medición se obtiene $99cm$,

Errores Absolutos y Relativos.

Una debilidad de esta definición es que la magnitud del error verdadero depende de la escala.

- Por ejemplo podemos medir una barra en centímetros o en metros. Si la longitud exacta de la barra es $1m$ y por la medición se obtiene $99cm$,
  1. $E_a = 100 - 99 = 1$, si usamos centímetros.

Errores Absolutos y Relativos.

Una debilidad de esta definición es que la magnitud del error verdadero depende de la escala.

- Por ejemplo podemos medir una barra en centímetros o en metros. Si la longitud exacta de la barra es $1m$ y por la medición se obtiene $99cm$,
    1. $E_a = 100 - 99 = 1$, si usamos centímetros.
    2. $E_a = 1.00 - 0.99 = 0.01$, si usamos metros.

Errores Absolutos y Relativos.

Una debilidad de esta definición es que la magnitud del error verdadero depende de la escala.

- Por ejemplo podemos medir una barra en centímetros o en metros. Si la longitud exacta de la barra es $1m$ y por la medición se obtiene $99cm$,
    1. $E_a = 100 - 99 = 1$, si usamos centímetros.
    2. $E_a = 1.00 - 0.99 = 0.01$, si usamos metros.

Esta es la razón por la que se define el error relativo.

Errores Absolutos y Relativos.

Al cociente entre el error absoluto E_a y el valor real x se le denomina error relativo y se denota por E_r. Se expresa también en valor absoluto, es decir:

Errores Absolutos y Relativos.

Al cociente entre el error absoluto E_a y el valor real x se le denomina error relativo y se denota por E_r. Se expresa también en valor absoluto, es decir:

$$E_r = \frac{|E_a|}{|x|} = \frac{|x - \tilde{x}|}{|x|}$$

Errores Absolutos y Relativos.

- Es preferible trabajar con errores relativos pues se toma en cuenta las magnitudes de los números con los que se está trabajando.

Errores Absolutos y Relativos.

- Es preferible trabajar con errores relativos pues se toma en cuenta las magnitudes de los números con los que se está trabajando.
- El uso del error absoluto tiene sentido sólamente si se tiene información a priori de estas magnitudes.

Errores Absolutos y Relativos.

- Es preferible trabajar con errores relativos pues se toma en cuenta las magnitudes de los números con los que se está trabajando.
- El uso del error absoluto tiene sentido sólamente si se tiene información a priori de estas magnitudes.
- De este modo un valor aproximado puede ser expresado de la siguiente manera en función del error relativo cometido y el valor real:

$$E_r = \frac{E_a}{x} = \frac{\tilde{x} - x}{x} \Rightarrow xE_r = \tilde{x} - x \Rightarrow x + xE_r = \tilde{x} \Rightarrow \tilde{x} = x(1 + E_r)$$

Cifras Significativas.

Se dice que el número $\tilde{x}$ aproxima al número x con t dígitos (o cifras) significativas, si t es el número más grande no negativo para el cual:

$$E_r < 0.5 \times 10^{-t} \Rightarrow \frac{|x - \tilde{x}|}{|x|} < 0.5 \times 10^{-t}$$

Cifras Significativas.

Se dice que el número $\tilde{x}$ aproxima al número x con t dígitos (o cifras) significativas, si t es el número más grande no negativo para el cual:

$$E_r < 0.5 \times 10^{-t} \Rightarrow \frac{|x - \tilde{x}|}{|x|} < 0.5 \times 10^{-t}$$

Ejemplo:

Sea $\tilde{x} = 3.1416$ una aproximación al valor π, y $x = 3.1415927$ una mejor aproximación.

$$\begin{aligned} E_a &= |x - \tilde{x}| = |3.1415927 - 3.1416| = 0.0000073 \\ E_r &= \frac{E_a}{|x|} = \frac{0.0000073}{3.1415927} = 0.0000023237 \\ t &< -\frac{\ln(2E_r)}{\ln(10)} = 5.3329 \end{aligned}$$

Cifras Significativas.

Ejemplo: Hallar el rango de aproximaciones con 4 cifras significativas para $x = 1000$

Cifras Significativas.

Ejemplo: Hallar el rango de aproximaciones con 4 cifras significativas para $x = 1000$

$$\frac{|1000 - \tilde{x}|}{|1000|} < 5 \times 10^{-4} \Rightarrow |1000 - \tilde{x}| < 5 \times 10^{-1} \Rightarrow |1000 - \tilde{x}| < 0,5$$

$$-0.5 < 1000 - \tilde{x} < 0.5 \Rightarrow 999.5 < \tilde{x} < 1000.5$$

Rango $= (999.5; 1000.5)$

Cifras Significativas.

Ejemplo: Hallar el rango de aproximaciones con 4 cifras significativas para $x = 1000$

$$\frac{|1000 - \tilde{x}|}{|1000|} < 5 \times 10^{-4} \Rightarrow |1000 - \tilde{x}| < 5 \times 10^{-1} \Rightarrow |1000 - \tilde{x}| < 0,5$$

$$-0.5 < 1000 - \tilde{x} < 0.5 \Rightarrow 999.5 < \tilde{x} < 1000.5$$

$$\text{Rango} \ = (999.5; 1000.5)$$

Observación

Las cifras significativas dan una idea de la exactitud en términos del Error Relativo.

Sistemas de numeración en base β

Un número N, en un sistema de numeración posicional, se representa como:

Sistemas de numeración en base β

Un número N, en un sistema de numeración posicional, se representa como:

$$N = (a_n a_{n-1} a_{n-2} \dots a_2 a_1 a_0)_\beta = \sum_{k=0}^{n} a_k \times \beta^k$$

donde:

- β: base o raíz del sistema numérico.
- a_k: dígitos o símbolos del sistema numérico. $0 \le a_k < \beta$

Un número N con parte decimal, en un sistema de numeración posicional, se representa como:

Un número N con parte decimal, en un sistema de numeración posicional, se representa como:

$$N = (a_n a_{n-1} \ldots a_2 a_1 a_0 . b_1 b_2 \ldots b_m)_\beta = \sum_{k=0}^{n} a_k \times \beta^k + \sum_{k=1}^{m} b_k \times \beta^{-k}$$

donde:

- β: base o raíz del sistema numérico.
- a_k, b_k: dígitos o símbolos del sistema numérico.
  $0 \leq a_k, b_k < \beta$
- n: número de dígitos enteros.
- m: número de dígitos fraccionarios.

Un número N con parte decimal, en un sistema de numeración posicional, se representa como:

$$N = (a_n a_{n-1} \ldots a_2 a_1 a_0 . b_1 b_2 \ldots b_m)_\beta = \sum_{k=0}^{n} a_k \times \beta^k + \sum_{k=1}^{m} b_k \times \beta^{-k}$$

donde:

- β: base o raíz del sistema numérico.
- a_k, b_k: dígitos o símbolos del sistema numérico.
  $0 \leq a_k, b_k < \beta$
- n: número de dígitos enteros.
- m: número de dígitos fraccionarios.
- $x_{10} = 27.5_{10} = 2 \times 10^1 + 7 \times 10^0 + 5 \times 10^{-1}$

Un número N con parte decimal, en un sistema de numeración posicional, se representa como:

$$N = (a_n a_{n-1} \ldots a_2 a_1 a_0 . b_1 b_2 \ldots b_m)_\beta = \sum_{k=0}^{n} a_k \times \beta^k + \sum_{k=1}^{m} b_k \times \beta^{-k}$$

donde:

- β: base o raíz del sistema numérico.
- a_k, b_k: dígitos o símbolos del sistema numérico. $0 \leq a_k, b_k < \beta$
- n: número de dígitos enteros.
- m: número de dígitos fraccionarios.
- $x_{10} = 27.5_{10} = 2 \times 10^1 + 7 \times 10^0 + 5 \times 10^{-1}$
- $x_2 = 101.01 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2}$

Sistemas de numeración en base β

La conversión a decimales es, por definición:

$$(a_n a_{n-1} \ldots a_1 a_0 . b_1 b_2 \ldots)_\beta = \sum_{i=0}^{n} a_i \beta^i + \sum_{i=1} b_i \beta^{-i} \qquad (1)$$

El sistema natural de numeración digital es el binario (base 2), utilizando sólo los dígitos 0 y 1.

$$101100.11_2 = 1 \times 2^5 + 1 \times 2^3 + 1 \times 2^2 + 1 \times 2^{-1} + 1 \times 2^{-2}$$

$$= 32 + 8 + 4 + 0.5 + 0.25 = 44.75$$

Conversión de base decimal a base β

La conversión de base decimal a base β se basa en el hecho de que, acudiendo a la definición (1)
se puede ver que:

$$(a_n a_{n-1} \ldots a_1 a_0.b_1 b_2 \ldots)_\beta \times \beta = (a_n a_{n-1} \ldots a_1 a_0 b_1.b_2 \ldots)_\beta$$
$$(a_n a_{n-1} \ldots a_1 a_0.b_1 b_2 \ldots)_\beta \times \beta^{-1} = (a_n a_{n-1} \ldots a_1.a_0 b_1 b_2 \ldots)_\beta \quad (2)$$

Conversión de base decimal a base β

La conversión de base decimal a base β se basa en el hecho de que, acudiendo a la definición (1)
se puede ver que:

$$(a_n a_{n-1} \ldots a_1 a_0 . b_1 b_2 \ldots)_\beta \times \beta = (a_n a_{n-1} \ldots a_1 a_0 b_1 . b_2 \ldots)_\beta$$
$$(a_n a_{n-1} \ldots a_1 a_0 . b_1 b_2 \ldots)_\beta \times \beta^{-1} = (a_n a_{n-1} \ldots a_1 . a_0 b_1 b_2 \ldots)_\beta \quad (2)$$

De (2) se deduce que:

$$(a_n \ldots a_0)_\beta \times \beta^{-1} = (a_n \ldots a_1)_\beta + (.a_0)_\beta$$

es decir, que

$$(a_n \ldots a_0)_\beta = (a_n \ldots a_1)_\beta \times \beta + (.a_0)_\beta \times \beta = (a_n \ldots a_1)_\beta \times \beta + (a_0)_\beta$$

Conversión de base decimal a base β

Así mismo de (2) se tiene que

$$(.b_1 b_2 \dots b_k)_\beta \times \beta = (b_1)_\beta + (.b_2 \dots b_k)_\beta$$

Conversión de base decimal a base β

Así mismo de (2) se tiene que

$$(.b_1 b_2 \dots b_k)_\beta \times \beta = (b_1)_\beta + (.b_2 \dots b_k)_\beta$$

$$N_{10} = (.625)_{10}$$

Conversión de base decimal a base β

Así mismo de (2) se tiene que

$$(.b_1 b_2 \ldots b_k)_\beta \times \beta = (b_1)_\beta + (.b_2 \ldots b_k)_\beta$$

$$\begin{aligned}
N_{10} &= (.625)_{10} \\
(.625)_{10} \times 2 &= (0.b_1 b_2 \ldots)_2 \times 2 = (b_1)_2 + (.b_2 b_3 \ldots)_2 \\
(1.25)_{10} &= (1.0 + 0.25)_{10} = (b_1)_2 + (.b_2 b_3 \ldots)_2 \Rightarrow b_1 = 1
\end{aligned}$$

Conversión de base decimal a base β

Así mismo de (2) se tiene que

$$(.b_1 b_2 \dots b_k)_\beta \times \beta = (b_1)_\beta + (.b_2 \dots b_k)_\beta$$

$$\begin{aligned}
N_{10} &= (.625)_{10} \\
(.625)_{10} \times 2 &= (0.b_1 b_2 \dots)_2 \times 2 = (b_1)_2 + (.b_2 b_3 \dots)_2 \\
(1.25)_{10} &= (1.0 + 0.25)_{10} = (b_1)_2 + (.b_2 b_3 \dots)_2 \Rightarrow b_1 = 1 \\
(.25)_{10} \times 2 &= (0.b_2 b_3 \dots)_2 \times 2 = (b_2)_2 + (.b_3 b_4 \dots)_2 \\
(0.5)_{10} &= (0.0 + 0.5)_{10} = (b_2)_2 + (.b_3 b_4 \dots)_2 \Rightarrow b_2 = 0
\end{aligned}$$

Conversión de base decimal a base β

Así mismo de (2) se tiene que

$$(.b_1 b_2 \ldots b_k)_\beta \times \beta = (b_1)_\beta + (.b_2 \ldots b_k)_\beta$$

$$\begin{aligned}
N_{10} &= (.625)_{10} \\
(.625)_{10} \times 2 &= (0.b_1 b_2 \ldots)_2 \times 2 = (b_1)_2 + (.b_2 b_3 \ldots)_2 \\
(1.25)_{10} &= (1.0 + 0.25)_{10} = (b_1)_2 + (.b_2 b_3 \ldots)_2 \Rightarrow b_1 = 1 \\
(.25)_{10} \times 2 &= (0.b_2 b_3 \ldots)_2 \times 2 = (b_2)_2 + (.b_3 b_4 \ldots)_2 \\
(0.5)_{10} &= (0.0 + 0.5)_{10} = (b_2)_2 + (.b_3 b_4 \ldots)_2 \Rightarrow b_2 = 0 \\
(.5)_{10} \times 2 &= (0.b_3 b_4 \ldots)_2 \times 2 = (b_3)_2 + (.b_4 b_5 \ldots)_2 \\
(1.0)_{10} &= (1.0 + 0.0)_{10} = (b_3)_2 + (.b_4 b_5 \ldots)_2 \Rightarrow b_3 = 1
\end{aligned}$$

Conversión de base decimal a base β

Así mismo de (2) se tiene que

$$(.b_1 b_2 \ldots b_k)_\beta \times \beta = (b_1)_\beta + (.b_2 \ldots b_k)_\beta$$

$$\begin{aligned}
N_{10} &= (.625)_{10} \\
(.625)_{10} \times 2 &= (0.b_1 b_2 \ldots)_2 \times 2 = (b_1)_2 + (.b_2 b_3 \ldots)_2 \\
(1.25)_{10} &= (1.0 + 0.25)_{10} = (b_1)_2 + (.b_2 b_3 \ldots)_2 \Rightarrow b_1 = 1 \\
(.25)_{10} \times 2 &= (0.b_2 b_3 \ldots)_2 \times 2 = (b_2)_2 + (.b_3 b_4 \ldots)_2 \\
(0.5)_{10} &= (0.0 + 0.5)_{10} = (b_2)_2 + (.b_3 b_4 \ldots)_2 \Rightarrow b_2 = 0 \\
(.5)_{10} \times 2 &= (0.b_3 b_4 \ldots)_2 \times 2 = (b_3)_2 + (.b_4 b_5 \ldots)_2 \\
(1.0)_{10} &= (1.0 + 0.0)_{10} = (b_3)_2 + (.b_4 b_5 \ldots)_2 \Rightarrow b_3 = 1 \\
(0.625)_{10} &= (0.101)_2
\end{aligned}$$

Dado un número fraccional cualquiera, el hecho de que su representación sea finita o infinita depende exclusivamente de la base utilizada en la representación.

Dado un número fraccional cualquiera, el hecho de que su representación sea finita o infinita depende exclusivamente de la base utilizada en la representación.

- La fracción $(0.1)_{10}$ no posee representación finita en base 2

$$(0.1)_{10} = 0.00011001100110011\dots$$

Dado un número fraccional cualquiera, el hecho de que su representación sea finita o infinita depende exclusivamente de la base utilizada en la representación.

- La fracción $(0.1)_{10}$ no posee representación finita en base 2

$$(0.1)_{10} = 0.00011001100110011\ldots$$

- La fracción $\frac{1}{3}$ que en base decimal tiene representación infinita periódica $0.\overline{3}$, en base ternaria (3) tendrá la representación finita 0.1.

Aproximación de Números.

Hay dos formas de aproximar un número:

- Por truncamiento.

Aproximación de Números.

Hay dos formas de aproximar un número:

- Por truncamiento.
- Por redondeo correcto.

Aproximación de Números.

Sea $x = a_n \ldots a_0.b_1 b_2 \ldots \in \mathbb{R}$ (En cualquier base), para redondear hasta el t-ésimo decimal:

- Por truncamiento

$$\tilde{x} = a_n \ldots a_0.b_1 b_2 \ldots b_t$$

Aproximación de Números.

Sea $x = a_n \dots a_0.b_1 b_2 \dots \in \mathbb{R}$ (En cualquier base), para redondear hasta el t-ésimo decimal:

- Por truncamiento

$$\tilde{x} = a_n \dots a_0.b_1 b_2 \dots b_t$$

- Por redondeo correcto

$$\tilde{x} \begin{cases} x - (0.b_{t+1} \dots) \times \beta^{-t} + \beta^{-t} & \text{si } (0.b_{t+1} \dots) \times \beta^{-t} \geq (1/2) \times \beta^{-t} \\ a_n \dots a_0.b_1 b_2 \dots b_t & \text{si } (0.b_{t+1} \dots) \times \beta^{-t} < (1/2) \times \beta^{-t} \end{cases}$$

Aproximación de Números.

Las cotas para el error absoluto y relativo vienen dadas por las siguientes expresiones:

- Por truncamiento

Aproximación de Números.

Las cotas para el error absoluto y relativo vienen dadas por las siguientes expresiones:

- Por truncamiento

$$E_a \leq \beta^{-t}$$

Aproximación de Números.

Las cotas para el error absoluto y relativo vienen dadas por las siguientes expresiones:

- Por truncamiento

$$E_a \leq \beta^{-t}$$

$$E_r \leq \beta^{-t+1}$$

Aproximación de Números.

Las cotas para el error absoluto y relativo vienen dadas por las siguientes expresiones:

- Por truncamiento

$$E_a \leq \beta^{-t}$$

$$E_r \leq \beta^{-t+1}$$

- Por redondeo correcto

Aproximación de Números.

Las cotas para el error absoluto y relativo vienen dadas por las siguientes expresiones:

- Por truncamiento

$$E_a \leq \beta^{-t}$$

$$E_r \leq \beta^{-t+1}$$

- Por redondeo correcto

$$E_a \leq \frac{1}{2} \times \beta^{-t}$$

Aproximación de Números.

Las cotas para el error absoluto y relativo vienen dadas por las siguientes expresiones:

- Por truncamiento

$$E_a \leq \beta^{-t}$$

$$E_r \leq \beta^{-t+1}$$

- Por redondeo correcto

$$E_a \leq \frac{1}{2} \times \beta^{-t}$$

$$E_r \leq \frac{1}{2} \times \beta^{-t+1}$$

Ejercicios:

1. Calcule el error absoluto, el error relativo y el número de cifras significativas en aproximaciones de $p = \pi$ mediante $\tilde{p}$:
   - $\tilde{p} = 22/7$ (en Antiguo Egipto, siglo XXVI a. C.)
   - $\tilde{p} = 223/71$ (Arquímedes, Antigua Grecia, siglo III a. C.)
   - $\tilde{p} = 3.14159$ (Liu Hui, China, año 265).
   - $\tilde{p} = 355/113$ (Zu Chongzhi, China, año 480).
2. Determine el mayor intervalo en que debe estar $\tilde{p}$ para aproximar p con un error relativo de a lo sumo 10^{-4} para cada valor de p:
   - $p = \pi$
   - $p = \sqrt[3]{7}$

Forma Normalizada de un Número

- La representación del sistema de números reales en un computador basa su idea en la conocida notación científica.

Forma Normalizada de un Número

- La representación del sistema de números reales en un computador basa su idea en la conocida notación científica.
- La notación científica permite representar números reales sobre un amplio rango de valores con sólo unos pocos dígitos.

Forma Normalizada de un Número

- La representación del sistema de números reales en un computador basa su idea en la conocida notación científica.
- La notación científica permite representar números reales sobre un amplio rango de valores con sólo unos pocos dígitos.
- Así 976000000000000 se representa como 9.76×10^{14} y 0.0000000000000976 como 9.76×10^{-14}.

Forma Normalizada de un Número

- La representación del sistema de números reales en un computador basa su idea en la conocida notación científica.
- La notación científica permite representar números reales sobre un amplio rango de valores con sólo unos pocos dígitos.
- Así 976000000000000 se representa como 9.76×10^{14} y 0.0000000000000976 como 9.76×10^{-14}.
- En esta notación el punto decimal se mueve dinámicamente a una posición conveniente y se utiliza el exponente de 10 para registrar la posición del punto decimal.

Forma Normalizada de un Número

- La representación del sistema de números reales en un computador basa su idea en la conocida notación científica.
- La notación científica permite representar números reales sobre un amplio rango de valores con sólo unos pocos dígitos.
- Así 976000000000000 se representa como 9.76×10^{14} y 0.0000000000000976 como 9.76×10^{-14}.
- En esta notación el punto decimal se mueve dinámicamente a una posición conveniente y se utiliza el exponente de 10 para registrar la posición del punto decimal.
- En particular, todo número real no nulo puede ser escrito en forma única en la notación científica normalizada.

Forma Normalizada de un Número

Un número del computador o de punto flotante, distinto de cero, se describe matemáticamente en la forma:

$$\sigma \times (0.a_1 a_2 \ldots a_t)_\beta \times \beta^e$$

donde

Forma Normalizada de un Número

Un número del computador o de punto flotante, distinto de cero, se describe matemáticamente en la forma:

$$\sigma \times (0.a_1 a_2 \ldots a_t)_\beta \times \beta^e$$

donde

- $\sigma = +1$ o $\sigma = -1$ es el signo del número.

Forma Normalizada de un Número

Un número del computador o de punto flotante, distinto de cero, se describe matemáticamente en la forma:

$$\sigma \times (0.a_1a_2\ldots a_t)_\beta \times \beta^e$$

donde

- $\sigma = +1$ o $\sigma = -1$ es el signo del número.
- β es un entero que denota la base del sistema numérico usado.

Forma Normalizada de un Número

Un número del computador o de punto flotante, distinto de cero, se describe matemáticamente en la forma:

$$\sigma \times (0.a_1a_2\dots a_t)_\beta \times \beta^e$$

donde

- $\sigma = +1$ o $\sigma = -1$ es el signo del número.
- β es un entero que denota la base del sistema numérico usado.
- a_i, $i = 1, 2, \dots, t$; es un entero con $0 \leq a_i \leq \beta - 1$, siendo $a_1 \neq 0$.

Forma Normalizada de un Número

Un número del computador o de punto flotante, distinto de cero, se describe matemáticamente en la forma:

$$\sigma \times (0.a_1 a_2 \ldots a_t)_\beta \times \beta^e$$

donde

- $\sigma = +1$ o $\sigma = -1$ es el signo del número.
- β es un entero que denota la base del sistema numérico usado.
- a_i, $i = 1, 2, \ldots, t$; es un entero con $0 \leq a_i \leq \beta - 1$, siendo $a_1 \neq 0$.
- e es un entero llamado el exponente, y es tal que $L \leq e \leq U$ para ciertos enteros L y U.

Forma Normalizada de un Número

De acuerdo con lo anterior un conjunto de punto flotante F queda caracterizado por cuatro parámetros:

- La base β.
- La precisión t.
- Los enteros L y U tales que $L \leq e \leq U$, donde e es el exponente.

Forma Normalizada de un Número

De acuerdo con lo anterior un conjunto de punto flotante F queda caracterizado por cuatro parámetros:

- La base β.
- La precisión t.
- Los enteros L y U tales que $L \leq e \leq U$, donde e es el exponente.

Una de las características de todo conjunto de punto flotante F es que es finito y tiene:

Forma Normalizada de un Número

De acuerdo con lo anterior un conjunto de punto flotante F queda caracterizado por cuatro parámetros:

- La base β.
- La precisión t.
- Los enteros L y U tales que $L \leq e \leq U$, donde e es el exponente.

Una de las características de todo conjunto de punto flotante F es que es finito y tiene:

$$2(\beta-1)\beta^{t-1}(U-L+1)+1$$

números diferentes (incluyendo el cero), y donde los distintos de cero están en forma normalizada.

Forma Normalizada de un Número

- Más aún, el conjunto $\mathbb{F}$ está acotado tanto superior como inferiormente, se tiene entonces que si se define:

Forma Normalizada de un Número

- Más aún, el conjunto $\mathbb{F}$ está acotado tanto superior como inferiormente, se tiene entonces que si se define:

$$F_L = (0.100\ldots0)_\beta \times \beta^L = \beta^{L-1}$$

como el número de punto flotante positivo más pequeño. Y

Forma Normalizada de un Número

- Más aún, el conjunto $\mathbb{F}$ está acotado tanto superior como inferiormente, se tiene entonces que si se define:

$$F_L = (0.100\ldots0)_\beta \times \beta^L = \beta^{L-1}$$

como el número de punto flotante positivo más pequeño. Y

$$F_U = (0.\gamma\gamma\ldots\gamma)_\beta \times \beta^U = (1-\beta^{-t})\beta^U \text{ con } \gamma = \beta - 1$$

como el número de punto flotante positivo más grande.

Forma Normalizada de un Número

- Más aún, el conjunto $\mathbb{F}$ está acotado tanto superior como inferiormente, se tiene entonces que si se define:

$$F_L = (0.100\ldots0)_\beta \times \beta^L = \beta^{L-1}$$

como el número de punto flotante positivo más pequeño. Y

$$F_U = (0.\gamma\gamma\ldots\gamma)_\beta \times \beta^U = (1-\beta^{-t})\beta^U \text{ con } \gamma = \beta - 1$$

como el número de punto flotante positivo más grande. Todo número $x \in \mathbb{F}$ satisface que:

$$F_L \leq |x| \leq F_U$$

Forma Normalizada de un Número

De las consideraciones anteriores se sigue, entonces, que en la recta de los números reales hay cuatro regiones excluidas para los números de $\mathbb{F}$, tal como se ilustra en la figura 1,

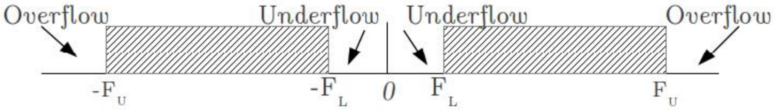

Figure: Números de punto flotante $\mathbb{F}(\beta, t, L, U)$.

Forma Normalizada de un Número

Sea el conjunto de punto flotante $\mathbb{F}$ con parámetros $\beta = 2$(Binario), $t = 3$, $L = -2$, $U = 2$. Tal conjunto F tiene

$$2(2-1)2^3 - 1(2-(-2)+1)+1 = 41$$

números diferentes (incluyendo el cero), en este caso, las mantisas serían $(0.100)_2$, $(0.101)_2$, $(0.110)_2$ y $(0.111)_2$ los cuales son la representación en base dos de los números reales $\frac{1}{2}$, $\frac{5}{8}$, $\frac{3}{4}$ y $\frac{7}{8}$ respectivamente, el total de números de máquina aparecen en la siguiente tabla

-2	-1	0	1	2
$(0.100)_2 \times 2^{-2}$	$(0.100)_2 \times 2^{-1}$	$(0.100)_2 \times 2^{0}$	$(0.100)_2 \times 2^{1}$	$(0.100)_2 \times 2^{2}$
$(0.101)_2 \times 2^{-2}$	$(0.101)_2 \times 2^{-1}$	$(0.101)_2 \times 2^{0}$	$(0.101)_2 \times 2^{1}$	$(0.101)_2 \times 2^{2}$
$(0.110)_2 \times 2^{-2}$	$(0.110)_2 \times 2^{-1}$	$(0.110)_2 \times 2^{0}$	$(0.110)_2 \times 2^{1}$	$(0.110)_2 \times 2^{2}$
$(0.111)_2 \times 2^{-2}$	$(0.111)_2 \times 2^{-1}$	$(0.111)_2 \times 2^{0}$	$(0.111)_2 \times 2^{1}$	$(0.111)_2 \times 2^{2}$

Table: Números binarios de $\mathbb{F}(2,3,-2,2)$

Forma Normalizada de un Número

- Los 41 números de máquina de este conjunto son los siguientes:

$$0, \pm\frac{4}{32}, \pm\frac{5}{32}, \pm\frac{6}{32}, \pm\frac{7}{32}, \pm\frac{8}{32}, \pm\frac{10}{32}, \pm\frac{12}{32}, \pm\frac{14}{32}, \pm\frac{16}{32}, \pm\frac{20}{32}, \pm\frac{24}{32}, \pm$$
$$\pm\frac{32}{32}, \pm\frac{40}{32}, \pm\frac{48}{32}, \pm\frac{56}{32}, \pm\frac{64}{32}, \pm\frac{80}{32}, \pm\frac{96}{32}, \pm\frac{112}{32}$$

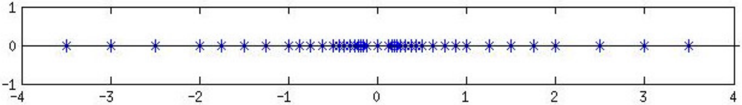

Figure: Números de punto flotante $\mathbb{F}(2, 3, -2, 2)$.

Forma Normalizada de un Número

- La combinación aritmética usual $+$, $-$, $\times$, $\div$ de dos números de punto flotante no siempre produce un número de punto flotante.

- Las operaciones aritméticas que realiza un computador no corresponden de forma exacta con las operaciones usuales. El estudio de lo que ocurre realmente es difícil de realizar y en todo caso depende de la máquina que se esté utilizando.

Forma Normalizada de un Número

- La combinación aritmética usual $+$, $-$, $\times$, $\div$ de dos números de punto flotante no siempre produce un número de punto flotante.
- Supongamos que $fl(x)$, $fl(y) \in \mathbb{F}$. Veamos, como ejemplo, que la suma usual $fl(x) + fl(y)$ no necesariamente será un número en $\mathbb{F}$. Sea el conjunto $\mathbb{F}$ dado en el ejemplo: $fl(x) = 5/32 \in \mathbb{F}$, $fl(y) = 48/32 \in \mathbb{F}$, sin embargo $fl(x) + fl(y) = 5/32 + 48/32 = 53/32 \notin \mathbb{F}$.
- Las operaciones aritméticas que realiza un computador no corresponden de forma exacta con las operaciones usuales. El estudio de lo que ocurre realmente es difícil de realizar y en todo caso depende de la máquina que se esté utilizando.

Forma Normalizada de un Número

- Denotando por $\oplus, \ominus, \otimes, \oslash$ las operaciones de suma, resta, multiplicación y división de la máquina. Se definen estas operaciones por:

$$\begin{aligned} x \oplus y &= fl(fl(x) + fl(y)) \\ x \ominus y &= fl(fl(x) - fl(y)) \\ x \otimes y &= fl(fl(x) \times fl(y)) \\ x \oslash y &= fl(fl(x)/fl(y)) \end{aligned}$$

Ejercicios:

Asumiendo $\beta = 10$, $t = 3$, $L = -3$, $U = 4$, y la aritmética es truncada. Obtener los valores de:

- $fl(0.00009)$
- $fl(3.146)$
- $fl(9996)$
- $fl((100.0 + 0.61) + 0.61)$ y $fl(100.0 + (0.61 + 0.61))$
- $fl(2.34 \times (5.67 + 8.90))$ y $fl((2.34 \times 5.67) + (2.34 \times 8.90))$

Unidad de Precisión o Redondeo

- En la representación en punto flotante con n dígitos en base β y exponente e, el error relativo en la representación de un número real x, $x \neq 0$ es estimado por:

$$\left|\frac{x - fl(x)}{x}\right| \leq \mu = \begin{cases} \beta^{1-n} & \text{si se trunca} \\ \frac{1}{2}\beta^{1-n} & \text{si se redondea} \end{cases}$$

Unidad de Precisión o Redondeo

- En la representación en punto flotante con n dígitos en base β y exponente e, el error relativo en la representación de un número real x, $x \neq 0$ es estimado por:

$$\left| \frac{x - fl(x)}{x} \right| \leq \mu = \begin{cases} \beta^{1-n} & \text{si se trunca} \\ \frac{1}{2}\beta^{1-n} & \text{si se redondea} \end{cases}$$

- El número μ (error de redondeo unitario) es una característica de la máquina, de su sistema operativo y de la manera en que efectúa los cálculos.

Unidad de Precisión o Redondeo

- En la representación en punto flotante con n dígitos en base β y exponente e, el error relativo en la representación de un número real x, $x \neq 0$ es estimado por:

$$\left| \frac{x - fl(x)}{x} \right| \leq \mu = \begin{cases} \beta^{1-n} & \text{si se trunca} \\ \frac{1}{2}\beta^{1-n} & \text{si se redondea} \end{cases}$$

- El número μ (error de redondeo unitario) es una característica de la máquina, de su sistema operativo y de la manera en que efectúa los cálculos.
- El épsilon de la máquina es importante porque caracteriza la precisión de la máquina, sirve además como criterio de parada de los algoritmos.

Unidad de Precisión o Redondeo

- Otra definición de $\mu \approx \epsilon$ (Épsilon de la máquina): ϵ es el más pequeño número positivo de la forma $\epsilon = 2^{-k}$ tal que:

$$1.0 + \epsilon \neq 1.0 \text{ en la máquina}$$

Unidad de Precisión o Redondeo

- Otra definición de $\mu \approx \epsilon$ (Épsilon de la máquina): ϵ es el más pequeño número positivo de la forma $\epsilon = 2^{-k}$ tal que:

$$1.0 + \epsilon \neq 1.0 \text{ en la máquina}$$

- $\epsilon = 2^{-n}$ donde n es la precisión de la máquina. Lo que sucede es que en la aritmética de la máquina llega un momento en que:

$$fl(1.0 + 2^{-k}) = 1.0$$

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$
2. para $k = 1, 2, \ldots, 100$ hacer

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$
2. para $k = 1, 2, \ldots, 100$ hacer
3. $\quad s = s * 0.5$

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$
2. para $k = 1, 2, \ldots, 100$ hacer
3. $\quad s = s * 0.5$
4. $\quad t = s + 1.0$

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$
2. para $k = 1, 2, \ldots, 100$ hacer
3. $\quad s = s * 0.5$
4. $\quad t = s + 1.0$
5. $\quad$ si $(t <= 1.0)$ entonces

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$
2. para $k = 1, 2, \ldots, 100$ hacer
3. $\quad s = s * 0.5$
4. $\quad t = s + 1.0$
5. $\quad$ si $(t <= 1.0)$ entonces
6. $\qquad s = 2.0 * s$

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$
2. para $k = 1, 2, \ldots, 100$ hacer
3. $\quad s = s * 0.5$
4. $\quad t = s + 1.0$
5. $\quad$ si $(t <= 1.0)$ entonces
6. $\qquad s = 2.0 * s$
7. $\qquad$ salida($k - 1$,s)

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$
2. para $k = 1, 2, \ldots, 100$ hacer
3. $\quad s = s * 0.5$
4. $\quad t = s + 1.0$
5. $\quad$ si $(t <= 1.0)$ entonces
6. $\quad\quad s = 2.0 * s$
7. $\quad\quad$ salida($k - 1$,s)
8. $\quad\quad$ parar

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$
2. para $k = 1, 2, \ldots, 100$ hacer
3. $\quad s = s * 0.5$
4. $\quad t = s + 1.0$
5. $\quad$ si $(t <= 1.0)$ entonces
6. $\qquad s = 2.0 * s$
7. $\qquad$ salida$(k - 1,s)$
8. $\qquad$ parar
9. $\quad$ fsi

Unidad de Presición o Redondeo.

Algoritmo para calcular ϵ

1. $s = 1$
2. para $k = 1, 2, \dots, 100$ hacer
3. $\quad s = s * 0.5$
4. $\quad t = s + 1.0$
5. $\quad$ si $(t <= 1.0)$ entonces
6. $\qquad s = 2.0 * s$
7. $\qquad$ salida$(k - 1, s)$
8. $\qquad$ parar
9. $\quad$ fsi
10. fpara

Propagación del Error.

- Es importante estudiar la propagación del error en los cálculos, ya que los errores se propagan y amplifican al realizar operaciones con dichos datos, hasta el punto de que puede suceder que el resultado carezca de significado.

Propagación del Error.

- Es importante estudiar la propagación del error en los cálculos, ya que los errores se propagan y amplifican al realizar operaciones con dichos datos, hasta el punto de que puede suceder que el resultado carezca de significado.
- Con el proposito de ilustrar esta situación, seguidamente se calcula la diferencia entre los números

$$a = 0.276435$$
$$b = 0.2756$$

Propagación del Error.

- Es importante estudiar la propagación del error en los cálculos, ya que los errores se propagan y amplifican al realizar operaciones con dichos datos, hasta el punto de que puede suceder que el resultado carezca de significado.
- Con el proposito de ilustrar esta situación, seguidamente se calcula la diferencia entre los números

$$a = 0.276435$$
$$b = 0.2756$$

Si los cálculos se realizan en base diez, punto flotante, redondeo correcto a tres dígitos de mantisa, los valores aproximados a dichos números y el error relativo cometido es

$$\tilde{a} = 0.276 \qquad |r_a| = 1.57 \times 10^{-3}$$
$$\tilde{b} = 0.276 \qquad |r_b| = 1.45 \times 10^{-3}$$

Propagación del Error

- Si ahora se calcula la diferencia entre los valores exactos y la diferencia entre los aproximados se obtiene

$$a - b = 0.000835$$
$$\tilde{a} - \tilde{b} = 0.0$$

Propagación del Error

- Si ahora se calcula la diferencia entre los valores exactos y la diferencia entre los aproximados se obtiene

$$a - b = 0.000835$$
$$\tilde{a} - \tilde{b} = 0.0$$

- Debe observa que el error relativo de la diferencia aproximada es del 100%. Este ejemplo, extraordinariamente sencillo, pone de manifiesto como el error de redondeo de los datos se ha amplificado al realizar una única operación, hasta generar un resultado carente de significado.

Propagación del Error en la Suma

Denotando por x e y los valores exactos de dos números y por $\tilde{x}$ e $\tilde{y}$ sus valores aproximados. Así mismo, los errores absolutos y relativos de estas cantidades se denotarán por E_x, E_y, r_x, r_y, respectivamente. Si se representa por $s = x + y$ al valor exacto de la suma y por $\tilde{s} = \tilde{x} + \tilde{y}$ su valor aproximado, entonces el error absoluto de la suma es

$$E_s = s - \tilde{s} = (x + y) - (\tilde{x} + \tilde{y}) = E_x + E_y$$

El error relativo vale

$$r_s = \frac{E_s}{s} = \frac{E_x + E_y}{x + y} = \frac{x}{x + y} r_x + \frac{y}{x + y} r_y$$

Propagación del Erroren la Resta

La deducción para la propagación del error mediante la resta es muy parecida a la anterior. Si se representa por $r = x - y$ al valor exacto de la resta y por $\tilde{r} = \tilde{x} - \tilde{y}$ su valor aproximado, entonces el error absoluto es

$$E_r = r - \tilde{r} = (x - y) - (\tilde{x} - \tilde{y}) = E_x - E_y$$

El error relativo vale

$$r_r = \frac{E_r}{r} = \frac{E_x - E_y}{x - y} = \frac{x}{x - y} r_x - \frac{y}{x - y} r_y$$

Propagación del Error en el Producto

Si se representa el producto de dos números exactos mediante $p = xy$ y el valor aproximado del producto por $\tilde{p} = \tilde{x}\tilde{y}$, el error absoluto del producto se puede calcular como

$$\begin{aligned} E_p &= p - \tilde{p} = (xy) - (\tilde{x}\tilde{y}) = xy - (x - E_x)(y - E_y) \\ &= xE_y + yE_x - E_xE_y \approx xE_y + yE_x \end{aligned}$$

El error relativo vale

$$r_p = \frac{E_p}{p} = \frac{xE_y + yE_x - ExEy}{xy} = r_x + r_y + r_xr_y \approx r_x + r_y$$

Propagación del Error en la División

Si se representa el cociente de dos números exactos mediante $d = x/y$ y el valor aproximado del producto por $\tilde{d} = \tilde{x}/\tilde{y}$, el error absoluto del cociente se puede calcular como

$$\begin{aligned} E_d &= d - \tilde{d} = \frac{x}{y} - \frac{\tilde{x}}{\tilde{y}} = \frac{x}{y} - \frac{x - Ex}{y - Ey} \\ &= \frac{yE_x - xE_y}{y(y - Ey)} \approx \frac{yE_x - xE_y}{y^2} \end{aligned}$$

El error relativo vale

$$r_d = \frac{E_d}{d} = \frac{\frac{yE_x - xE_y}{y(y-Ey)}}{\frac{x}{y}} = \frac{yE_x - xE_y}{x(y - E_y)} = \frac{r_x - r_y}{1 - r_y} \approx r_x - r_y$$

Propagación del Error en una Función

Sea $z = f(x)$ la imagen mediante la función f del valor exacto de un número x y sea $\tilde{z} = f(\tilde{x})$ la imagen de su valor aproximado. Entonces, el error absoluto de la imagen es

$$\begin{aligned} E_z &= f(x) - f(\tilde{x}) = f(x) - f(x - E_x) \\ &= f(x) - \left[f(x) - f'(x)E_x + f''(x)\frac{E_x^2}{2!} - f'''(x)\frac{E_x^3}{3!} + \cdots \right] \\ &= f'(x)E_x - f''(x)\frac{E_x^2}{2!} + f'''(x)\frac{E_x^3}{3!} - \cdots \approx f'(x)E_x \end{aligned}$$

El error relativo vale

$$r_z = \frac{E_z}{z} = \frac{f'(x)E_x - f''(x)\frac{E_x^2}{2!} + f'''(x)\frac{E_x^3}{3!} - \cdots}{f(x)} \approx x\frac{f'(x)}{f(x)} r_x$$

Pérdida de Cifras Significativas.

- Este tipo de situación se da cuando algún cálculo envuelve la resta de dos cantidades similares ó en el cálculo de un número por medio de una sumatoria donde el resultado es menor que los términos de la sumatoria los cuales alternan en signo.

Pérdida de Cifras Significativas.

- Este tipo de situación se da cuando algún cálculo envuelve la resta de dos cantidades similares ó en el cálculo de un número por medio de una sumatoria donde el resultado es menor que los términos de la sumatoria los cuales alternan en signo.
- Por ejemplo: Sea $f(x) = \sqrt{x+1} - \sqrt{x}$ para x grande. Suponemos que trabajamos con seis cifras significativas. Veamos el caso de calcular $f(100)$.
  $\sqrt{100} = 10.000 \qquad \sqrt{101} = 10.0499$
  Calculamos $f(100) = \sqrt{101} - \sqrt{100} = 0.499000 \times 10^{-1}$. El valor exacto a seis cifras de $f(100)$ es 0.498756×10^{-1}

Pérdida de Cifras Significativas.

¿Cúal fué el problema? Se restaron cantidades similares. En este caso se puede evitar restando cantidades similares si se reescribe $f(x)$ (racionalizando) como:

$$f(x) = \frac{1}{\sqrt{x+1} + \sqrt{x}}$$

$$f(100) = \frac{1}{10.000 + 10.0499} = \frac{1}{20.0499} = 0.498756 \times 10^{-1}$$

Condicionamiento y Estabilidad.

- La condición de un problema o de una función es la medida de la sensibilidad de esa función a pequeños cambios en sus parámetros.

Condicionamiento y Estabilidad.

- La condición de un problema o de una función es la medida de la sensibilidad de esa función a pequeños cambios en sus parámetros.
- Una función está bien condicionada si pequeños cambios en los parámetros inducen solo un cambio pequeño en el comportamiento de la función; de otro modo, está mal condicionada.

Condicionamiento y Estabilidad.

- Supongamos que se quiere calcular la integral

$$E_n = \int_0^1 x^n e^{x-1} dx, \qquad n = 1, 2, \ldots$$

Condicionamiento y Estabilidad.

- Supongamos que se quiere calcular la integral

$$E_n = \int_0^1 x^n e^{x-1} dx, \qquad n = 1, 2, \ldots$$

- Usando integración por partes

$$\int_0^1 x^n e^{x-1} dx = x^n e^{x-1}\Big|_0^1 - \int_0^1 n x^{n-1} e^{x-1} dx$$

o bien

$$E_n = 1 - nE_{n-1}, \qquad n = 2, \ldots$$

Condicionamiento y Estabilidad.

- Donde $E_1 = e^{-1}$

Condicionamiento y Estabilidad.

- Donde $E_1 = e^{-1}$
- Supongamos que se comienza el proceso iterativo con $\tilde{E}_1 = E_1 + \varepsilon$ y efectuamos todas las operaciones aritméticas en forma exacta. Entonces,

$$
\begin{aligned}
\tilde{E}_2 &= 1 - 2\tilde{E}_1 = 1 - 2E_1 - 2\varepsilon = E_2 - 2\varepsilon \\
\tilde{E}_3 &= 1 - 3\tilde{E}_2 = 1 - 3E_2 + 6\varepsilon = E_3 + 3!\varepsilon \\
\tilde{E}_4 &= 1 - 4\tilde{E}_3 = 1 - 4E_3 - 24\varepsilon = E_4 - 4!\varepsilon \\
&\vdots \\
\tilde{E}_n &= E_n \pm n!\varepsilon
\end{aligned}
$$

Condicionamiento y Estabilidad.

- Un pequeño cambio en el valor inicial E_1 origina un gran cambio en los valores posteriores de E_n.

Condicionamiento y Estabilidad.

- Un pequeño cambio en el valor inicial E_1 origina un gran cambio en los valores posteriores de E_n.
- De hecho, en este caso, el efecto es devastador: puesto que los valores exactos de E_n decrecen conforme n aumenta, y a partir de cierto n el error será tan o más grande que el valor a estimar.

Condicionamiento y Estabilidad.

- Un pequeño cambio en el valor inicial E_1 origina un gran cambio en los valores posteriores de E_n.
- De hecho, en este caso, el efecto es devastador: puesto que los valores exactos de E_n decrecen conforme n aumenta, y a partir de cierto n el error será tan o más grande que el valor a estimar.
- Claramente, el algoritmo recursivo propuesto es numéricamente inestable.

Condicionamiento y Estabilidad.

- Para evitar el efecto devastador del error, se puede usar una fórmula recursiva que permita calcular aproximaciones a E_n a partir de aproximaciones a E_{n-1}.

Condicionamiento y Estabilidad.

- Para evitar el efecto devastador del error, se puede usar una fórmula recursiva que permita calcular aproximaciones a E_n a partir de aproximaciones a E_{n-1}.
- Considérese la fórmula recursiva escrita en forma inversa,

$$E_{n-1} = \frac{1 - E_n}{n}, \qquad n = \ldots, N, N-1, \ldots, 3, 2, 1$$

Condicionamiento y Estabilidad.

- Para evitar el efecto devastador del error, se puede usar una fórmula recursiva que permita calcular aproximaciones a E_n a partir de aproximaciones a E_{n-1}.
- Considérese la fórmula recursiva escrita en forma inversa,

$$E_{n-1} = \frac{1 - E_n}{n}, \qquad n = \ldots, N, N-1, \ldots, 3, 2, 1$$

- Si se conoce una aproximación $\tilde{E}_N$ a E_N para algún N, entonces la fórmula anterior permitiría calcular aproximaciones a $E_{N-1}, E_{N-2}, \ldots, E_1$.

Condicionamiento y Estabilidad.

- Para determinar la estabilidad del proceso recursivo considerese, en forma análoga a lo anterior, que $\tilde{E}_N = E_N + \varepsilon$, entonces

$$\tilde{E}_{N-1} = \frac{1-\tilde{E}_N}{N} = \frac{1-E_N}{N} - \frac{\varepsilon}{N} = E_{N-1} - \frac{\varepsilon}{N}$$

$$\tilde{E}_{N-2} = E_{N-2} + \frac{\varepsilon}{N(N-1)}$$

$$\vdots$$

$$\tilde{E}_1 = E_1 \pm \frac{\varepsilon}{N!}$$

Condicionamiento y Estabilidad.

- Se observa que el error introducido en E_N es rápidamente reducido en cada paso. Más aún, la magnitud del error tiende a ser mucho menor respecto del valor a estimar.

Condicionamiento y Estabilidad.

- Se observa que el error introducido en E_N es rápidamente reducido en cada paso. Más aún, la magnitud del error tiende a ser mucho menor respecto del valor a estimar.
- Esto último permite que, aún con una pobre aproximación $\tilde{E}_N$ para algún N suficientemente grande, se puede obtener una aproximación precisa para el valor E_n de interés.

Condicionamiento y Estabilidad.

- Se observa que el error introducido en E_N es rápidamente reducido en cada paso. Más aún, la magnitud del error tiende a ser mucho menor respecto del valor a estimar.
- Esto último permite que, aún con una pobre aproximación $\tilde{E}_N$ para algún N suficientemente grande, se puede obtener una aproximación precisa para el valor E_n de interés.
- Para obtener un valor inicial, nótese que:

$$E_n = \int_0^1 x^n e^{x-1} dx \leq \int_0^1 x^n dx = \left. \frac{x^{n+1}}{n+1} \right|_0^1 = \frac{1}{n+1}$$

Condicionamiento y Estabilidad.

- Por tanto, E_n tiende a cero cuando n tiende a ∞. Por ejemplo, si aproximamos E_{20} por cero y usamos este valor para empezar, como mucho cometemos un error de $\frac{1}{21}$.

Condicionamiento y Estabilidad.

- Por tanto, E_n tiende a cero cuando n tiende a ∞. Por ejemplo, si aproximamos E_{20} por cero y usamos este valor para empezar, como mucho cometemos un error de $\frac{1}{21}$.
- Este error es multiplicado por $\frac{1}{20}$ al calcular E_{19}, y el error en E_{19} es a lo más $\left(\frac{1}{21}\right)\left(\frac{1}{20}\right) \approx 0.0024$.

Condicionamiento y Estabilidad.

- Por tanto, E_n tiende a cero cuando n tiende a ∞. Por ejemplo, si aproximamos E_{20} por cero y usamos este valor para empezar, como mucho cometemos un error de $\frac{1}{21}$.
- Este error es multiplicado por $\frac{1}{20}$ al calcular E_{19}, y el error en E_{19} es a lo más $\left(\frac{1}{21}\right)\left(\frac{1}{20}\right) \approx 0.0024$.
- Al llegar a E_{15}, el error inicial ha sido reducido a menos de 4×10^{-8}, que ya es menor que el error de redondeo.